Capacity of the Discrete-Time Gaussian Channel with Intersymbol Interference

WALTER HIRT, MEMBER, IEEE, AND JAMES L. MASSEY, FELLOW, IEEE

Abstract —The discrete-time Gaussian channel with intersymbol interference (ISI) where the inputs are subject to a *per symbol* average-energy constraint is considered. The capacity of this channel is derived by means of a hypothetical channel model, called the N-circular Gaussian channel (NCGC), whose capacity is readily derived using the theory of the discrete Fourier transform. The results obtained for the NCGC are used further to prove that, in the limit of increasing block length, N, the capacity of the discrete-time Gaussian channel (DTGC) with ISI using a *per block* average-energy input constraint (N-block DTGC) is indeed also the capacity when using the *per symbol* average-energy constraint.

I. Introduction

There is presently much interest in the design and application of codes for channels with finite memory produced by linear filtering of the input digits [1]–[3]. This memory introduces intersymbol interference (ISI), which is generally considered to be an undesirable property of a pulse-amplitude modulated (PAM) digital communication channel [4]. There are situations, however, where it is sensible to introduce ISI intentionally. Partial-response schemes [5], for example, are designed to produce a controlled amount of ISI in the received signal in return for better spectral characteristics. What matters for the coding system is the equivalent discrete-time channel which is created by the actual transmission system. In this paper the capacity of such channels is of interest. In general, channel capacity can be defined and computed provided that the channel model includes 1) the *basic channel model* specifying the conditional probability for the output given a specified input, and 2) the *constraints* on channel usage. We introduce three related channel models for channels with ISI and define their respective capacities. These models differ in parts 1) or 2), or both, of the definition.

A. Discrete-Time Gaussian Channel with ISI

The well-known discrete-time model for the equivalent baseband channel of a PAM system with ISI and with zero-mean additive white Gaussian noise (AWGN) having one-sided power spectral density N_0 is the *basic channel model* of interest [4], [6], [7]. The real input sequence $\{x_k\}$ produces the real output sequence $\{y_k\}$ given by

$$y_k = \sum_{i=0}^{M} h_i x_{k-i} + w_k, \qquad -\infty < k < \infty \tag{1}$$

where the finite-length sequence $(h_0, h_1, \cdots, h_M)$, with $h_0 \neq 0$ and $h_M \neq 0$, is the unit-sample response of the equivalent channel filter. The transfer function of this filter is

$$H(\lambda) = \sum_{i=0}^{M} h_i e^{-ji\lambda}, \qquad j = \sqrt{-1} \tag{2}$$

and is periodic in λ with period 2π. The noise samples w_k are independent identically distributed (i.i.d.) Gaussian random variables with mean zero and variance $N_0/2$, i.e.,

$$E[w_k] = 0 \quad \text{and} \quad E[w_k w_i] = (N_0/2)\delta_{k-i} \tag{3}$$

where $E[\cdot]$ denotes expectation and where $\delta_0 = 1$ and $\delta_n = 0$ for $n \neq 0$. With the time unit taken as the interval between input symbols, M represents the memory of the channel. The channel is said to have ISI if and only if $M > 0$. Since M is finite, the energy of the unit-sample response $(h_0, h_1, \cdots, h_M)$ is also finite. For $m \le k < N$, we write (1) symbolically as

$$y[m, N-1] = x[m-M, N-1] * h[0, M] + w[m, N-1] \tag{4}$$

where the asterisk denotes the linear convolution operator and where here and hereafter we use the sequence notation $s[m,n] = (s_m, s_{m+1}, \cdots, s_n)$. Note that the subsequence $x[m-M, m-1]$ of the input sequence $x[m-M, N-1]$ represents the initial contents of the channel memory, i.e., the *channel state* at the time instant $k = m$ when we assume that we begin to observe the output.

We now require that the basic channel model (1) be used in a manner where the inputs individually satisfy the *constraint*

$$E[x_k^2] \le E_s, \qquad -\infty < k < \infty, \tag{5}$$

so that E_s is the maximum allowed *per symbol* average energy. We call the channel defined by both the basic channel model (1) and the input constraint (5) the *discrete-time Gaussian channel* (DTGC) with finite memory M. The DTGC is the natural channel model for the energy-constrained Gaussian channel with ISI, although the difficulty of dealing with constraint (5) has led to the

Manuscript received February 17, 1987; revised June 30, 1987. This paper was presented in part at the IEEE International Symposium on Information Theory, Brighton, England, June 24–28, 1985.
W. Hirt is with the IBM Research Division, IBM Zurich Research Laboratory, Säumerstrasse 4, 8803 Rüschlikon, Switzerland.
J. L. Massey is with the Institute for Signal and Information Processing, Swiss Federal Institute of Technology, 8092 Zurich, Switzerland.
IEEE Log Number 8821203.

0018-9448/88/0500-0380$01.00

more general use of a different channel model, namely, the block-energy-constrained channel described later.

It follows from Gallager [8, sections 4.6 and 5.9] that the appropriate definition of capacity for the DTGC is

$$C(E_s) = \lim_{N \to \infty} I_N(E_s) \tag{6a}$$

with

$$I_N(E_s) = \sup_{q_N} N^{-1} I(x[0, N-1]; y[0, N-1]) \tag{6b}$$

where the supremum of the average mutual information $I(\cdot;\cdot)$ is taken over all probability densities q_N for the sequence $x[0, N-1]$ satisfying the symbol-energy constraint (5), and where it is assumed that

$$x[-M, -1] = (0,0,\cdots,0). \tag{6c}$$

The choice of $x[-M, -1]$ in (6c) is made for convenience but has no influence on $C(E_s)$ since M is finite. Definition (6) is appropriate in the sense that a coding theorem and its converse can be proved to show that $C(E_s)$ is the upper limit of information rate (in bits per channel input symbol) such that arbitrarily reliable communication is possible over the DTGC.

B. N-Block DTGC

Our second channel model also includes the *basic channel model* of (1); however, the *constraint* on the inputs is now

$$\sum_{k=0}^{N-1} E\left[x_k^2\right] \le NE_s \tag{7}$$

where N is the *block length*. Thus NE_s is the maximum allowed *per block* average energy. We call the channel model as defined by (1) and (7) the *N-block DTGC*. Note that for any N, constraint (7) is weaker than constraint (5); constraint (5) always implies (7), but not conversely.

The capacity of the N-block DTGC is defined as

$$\hat{C}_N(E_s) = \sup_{q_N} N^{-1} I(x[0, N-1]; y[0, N-1]) \tag{8}$$

where the supremum is taken over all probability densities q_N for the sequence $x[0, N-1]$ satisfying the block-energy constraint (7) and where $x[-M, -1]$ is the same as in (6c). It is conventional to define the quantity

$$\hat{C}(E_s) = \lim_{N \to \infty} \hat{C}_N(E_s) \tag{9}$$

to be the "capacity" of the energy-constrained Gaussian channel with ISI [9]–[12], but it is important to note that $\hat{C}(E_s)$ is not an actual capacity because the N-block DTGC is (by definition) a *different* channel for each N. It seems intuitively obvious, however, that

$$C(E_s) = \hat{C}(E_s) \tag{10}$$

but (to our knowledge) this has never previously been proved. All one can claim at this point is that $C(E_s) \le \hat{C}(E_s)$ since the symbol-energy constraint (5) is stronger than the block-energy constraint (7). To prove the validity of (10), we introduce a new channel model whose capacity can be readily determined and used to relate $C(E_s)$ and $\hat{C}(E_s)$.

C. New Channel Model

We define a new channel model by modifying (1). The samples of the output sequence of the new *basic channel model* $\{\tilde{y}_k\}$ are determined by

$$\tilde{y}_k = \sum_{i=0}^{N-1} \tilde{h}_i x_{((k-i))} + w_k, \qquad 0 \le k < N \tag{11}$$

where $((\cdot))$ denotes addition modulo N and where $N > M$. Defining $\tilde{h}[0, N-1] = (h_0, h_1, \cdots, h_M, 0, 0, \cdots, 0)$ as the unit-sample response $h[0, M]$ extended with $N - M - 1$ zeros, we can write (11) symbolically as

$$\tilde{y}[0, N-1] = x[0, N-1] \circledast \tilde{h}[0, N-1] + w[0, N-1] \tag{12}$$

where $\circledast$ denotes the circular convolution operator. For our new channel model, the input *constraint* is

$$E\left[x_k^2\right] \le E_s, \qquad 0 \le k < N, \tag{13}$$

so that E_s is again the maximum allowed *per symbol* average energy. We call the channel model defined by (11) and (13) the *N-circular Gaussian channel* (NCGC).

The capacity of the NCGC is defined as

$$\tilde{C}_N(E_s) = \sup_{q_N} N^{-1} I(x[0, N-1]; \tilde{y}[0, N-1]) \tag{14}$$

where the supremum is taken over all probability densities q_N for the sequence $x[0, N-1]$ satisfying the symbol-energy constraint (13). Note that no need exists for an initializing input sequence since the output sequence $\tilde{y}[0, N-1]$ is completely determined from the input sequence $x[0, N-1]$ and the noise sequence $w[0, N-1]$. For the NCGC, we define the *asymptotic capacity*

$$\tilde{C}(E_s) = \lim_{N \to \infty} \tilde{C}_N(E_s), \tag{15}$$

which again is not itself a true capacity since the NCGC is (by definition) a *different* channel for each N.

D. Remarks

The capacity of the N-block DTGC, $\hat{C}_N(E_s)$, as well as its limit as $N \to \infty$, $\hat{C}(E_s)$, have been derived by Tsybakov [9], [10] and others [12], [13]. Tsybakov also treated the case where the channel memory is unbounded ($M = \infty$). In all cases, $\hat{C}_N(E_s)$ was obtained by solving an eigenvalue problem, and $\hat{C}(E_s)$ was found by invoking asymptotic properties of Toeplitz forms [11], [14].

We present an approach to finding the capacity of the DTGC, $C(E_s)$, based on the discrete Fourier transform (DFT). It seemed worthwhile giving a rigorous derivation of capacity for this important channel model using only the well-known theory of the DFT rather than the more specialized theory of Toeplitz forms. Moreover, our approach allows us to *prove* that the conventional "capacity"

$\hat{C}(E_s)$, defined as the limit as $N \to \infty$ of the capacity of the N-block DTGC $\hat{C}_N(E_s)$, is indeed also the capacity of the DTGC $C(E_s)$.

In Section II we state the main results of this paper, which are proved in the following sections. Section III contains an analysis of the new NCGC, and Section IV a proof of the fundamental relations between the channel models just introduced. The main results are proved in Section V, and Section VI provides a summary and concluding remarks.

II. Statement of Main Results

Theorem 1: The capacity of the NCGC (in *bits per channel input symbol* when logarithms are taken to the base 2) is given by

$$\tilde{C}_N(E_s) = (2N)^{-1} \sum_{i=0}^{N-1} \log\left[\max\left(\Theta|\tilde{H}_i|^2, 1\right)\right] \quad (16a)$$

where $\tilde{H}[0, N-1]$ is the DFT of $\tilde{h}[0, N-1]$, i.e.,

$$\tilde{H}_i = \sum_{m=0}^{N-1} \tilde{h}_m e^{-j2\pi im/N}, \qquad 0 \le i < N, \quad (16b)$$

and where the parameter Θ is the solution of

$$\sum_{\substack{i=0 \\ \tilde{H}_i \neq 0}}^{N-1} \max\left(\Theta - |\tilde{H}_i|^{-2}, 0\right) = 2NE_s/N_0. \quad (16c)$$

Moreover, for the capacity-achieving q_N, the components of the input sequence $x[0, N-1]$ are correlated Gaussian random variables with mean zero and covariances $\tilde{r}_n$, $0 \le n < N$, given by

$$\tilde{r}_n = E[x_{k+n}x_k] = N^{-1} \sum_{i=0}^{N-1} \epsilon_i \cos(2\pi ni/N), \qquad 0 \le k \le k+n < N \quad (17a)$$

where the components of the spectral input energy sequence $\epsilon[0, N-1]$ satisfy

$$\epsilon_i = \begin{cases} (N_0/2)\left(\Theta - |\tilde{H}_i|^{-2}\right), & \Theta|\tilde{H}_i|^2 > 1 \\ 0, & \text{otherwise.} \end{cases} \quad (17b)$$

In particular, capacity is achieved when equality holds in (13), i.e., when all inputs x_k, $0 \le k < N$, have the same average energy $E[x_k^2] = \tilde{r}_0 = E_s$. Equation (17a) implies that the covariance sequence $\tilde{r}[0, N-1]$ is the inverse DFT of the spectral energy sequence $\epsilon[0, N-1]$; conversely, $\epsilon[0, N-1]$ is the DFT of $\tilde{r}[0, N-1]$.

Corollary 1: The DTGC, the N-block DTGC, and the NCGC are asymptotically equivalent channel models in the sense that

$$C(E_s) = \hat{C}(E_s) = \tilde{C}(E_s). \quad (18)$$

Theorem 2: The capacity of the DTGC (in *bits per channel input symbol* when logarithms are taken to the base 2) is given by

$$C(E_s) = (2\pi)^{-1} \int_0^{\pi} \log\left[\max\left(\Theta|H(\lambda)|^2, 1\right)\right] d\lambda \quad (19a)$$

where $H(\lambda)$ is the channel transfer function given in (2) and where the parameter Θ is the solution of

$$\int_{\substack{0 \\ H(\lambda) \neq 0}}^{\pi} \max\left(\Theta - |H(\lambda)|^{-2}, 0\right) d\lambda = 2\pi E_s/N_0. \quad (19b)$$

Moreover, the capacity-achieving q_N, the inputs x_k, $-\infty < k < \infty$, are correlated Gaussian random variables with mean zero and covariances r_n, $-\infty < n < \infty$, given by

$$r_n = E[x_{k+n}x_k] = (\pi)^{-1} \int_0^{\pi} S_x(\lambda) \cos(n\lambda)\, d\lambda \quad (20a)$$

where the input power spectral density satisfies

$$S_x(\lambda) = \begin{cases} (N_0/2)\left(\Theta - |H(\lambda)|^{-2}\right), & \Theta|H(\lambda)|^2 > 1, |\lambda| \le \pi, \\ 0, & \text{otherwise.} \end{cases} \quad (20b)$$

In particular, capacity is achieved when equality holds in (5), i.e., when all inputs x_k, $-\infty < k < \infty$, have the same average energy $E[x_k^2] = r_0 = E_s$.

III. Analysis of the NCGC

The proof of Theorem 1 will be given by first considering a channel model with the basic channel model of the NCGC, given in (11), and the block-energy input constraint given in (7). We are thus interested in the quantity

$$\tilde{I}_N(E_s) = \sup_{q_N} N^{-1} I(x[0, N-1]; \tilde{y}[0, N-1]) \quad (21)$$

where the supremum is taken over all probability densities q_N for the sequence $x[0, N-1]$ satisfying the block-energy constraint (7). We shall then show that the optimizing q_N in (21) also satisfies the stronger symbol-energy constraint (13) so that $\tilde{C}_N(E_s) = \tilde{I}_N(E_s)$.

A. Derivation of $\tilde{I}_N(E_s)$

The DFT of a sequence $b[0, N-1] = (b_0, b_1, \cdots, b_{N-1})$ is the sequence $B[0, N-1] = (B_0, B_1, \cdots, B_{N-1})$ defined by

$$B_i = \sum_{k=0}^{N-1} b_k \Omega_N^{-ik}, \qquad 0 \le i < N \quad (22a)$$

where $\Omega_N = e^{j2\pi/N}$, $j = \sqrt{-1}$ [15, p. 100]. The inverse DFT is given by

$$b_k = N^{-1} \sum_{i=0}^{N-1} B_i \Omega_N^{ki}, \qquad 0 \le k < N. \quad (22b)$$

Taking the DFT on both sides of (12) yields

$$\tilde{Y}_i = \tilde{H}_i X_i + W_i, \qquad 0 \le i < N \quad (23)$$

as a result of the linearity and the circular convolution properties of the DFT [15, p. 110]. In (23), $\tilde{Y}_i$, $\tilde{H}_i$, X_i, and W_i are the components of DFT$\{\tilde{y}[0, N-1]\}$, DFT$\{\tilde{h}[0, N$

$-1]\}$, $\mathrm{DFT}\{x[0,N-1]\}$, and $\mathrm{DFT}\{w[0,N-1]\}$, respectively. In the following, let $L=\lfloor N/2\rfloor$, where $\lfloor\cdot\rfloor$ denotes the integer part of a real number.

For any sequence $B[0,N-1]=\mathrm{DFT}\{b[0,N-1]\}$ where $b[0,N-1]$ is real, $B_i=B^*_{N-i}$, $1\le i<N$ [15, p. 110] (here $*$ denotes complex conjugate). Note that B_0 is real for N even or odd, and B_L is real for N even. Therefore, knowledge of the components B_i, $0\le i\le L$, is sufficient to reconstruct the real sequence $b[0,N-1]$, and no information is lost by discarding the components B_i, $L<i<N$. The remaining complex components B_i, $0\le i\le L$, may be further transformed according to

$$B'_i=\begin{cases} B_i=B_i^R, & i=\begin{cases}0, & N \text{ odd}\\ 0 \text{ or } L, & N\text{ even}\end{cases}\\ \sqrt{2}\,B_i^R, & 1\le i\le\begin{cases}L, & N\text{ odd}\\ L-1, & N\text{ even}\end{cases}\\ \sqrt{2}\,B^I_{N-i}, & L<i<N\end{cases} \tag{24a}$$

where $B_i^R=\mathrm{Re}(B_i)$ and $B_i^I=\mathrm{Im}(B_i)$. The obvious inverse of (24a) is

$$B_i=\begin{cases} B'_i, & i=\begin{cases}0, & N\text{ odd}\\ 0\text{ or } L, & N\text{ even}\end{cases}\\ (B'_i+jB'_{N-i})/\sqrt{2}, & 1\le i\le\begin{cases}L, & N\text{ odd}\\ L-1, & N\text{ even}\end{cases}\\ B^*_{N-i}, & L<i<N.\end{cases} \tag{24b}$$

To obtain an equivalent representation of (11), we divide both sides of (23) by the complex constant $\tilde{H}_i$ (assuming temporarily that $\tilde{H}_i\ne 0$ for all i) and transform the resulting first $L+1$ components with (24a). The equivalent form of (11) in the transform domain is now

$$Y'_i=X'_i+V'_i,\qquad 0\le i<N \tag{25}$$

where the X'_i and V'_i were obtained from (24a) with $B_i\equiv X_i$ and $B_i\equiv W_i/\tilde{H}_i$, respectively, and where $\tilde{H}_i\ne 0$, $0\le i\le L$. Later, it will become clear that this restriction on $\tilde{H}_i$ is not necessary.

It will be useful to combine the transforms defined in (22) and (24). The relations between the real time-domain variables b_k, $0\le k<N$, and the real transform-domain variables B'_i, $0\le i<N$, may be written as the transform pair

$$B'_i=\begin{cases} d_i\sum_{k=0}^{N-1} b_k\cos(2\pi ik/N), & 0\le i\le L\\ d_i\sum_{k=1}^{N-1} b_k\sin(2\pi ik/N), & L<i<N\end{cases} \tag{26a}$$

$$b_k=N^{-1}\left[\sum_{i=0}^{L} d_iB'_i\cos(2\pi ki/N)+\sum_{i=L+1}^{N-1} d_iB'_i\sin(2\pi ki/N)\right],\qquad 0\le k<N \tag{26b}$$

where

$$d_i=\begin{cases}1, & i=\begin{cases}0, & N\text{ odd}\\ 0\text{ or } L, & N\text{ even}\end{cases}\\ \sqrt{2}, & \text{otherwise.}\end{cases} \tag{26c}$$

Transform (26) is a form of the real discrete Fourier transform (RDFT) [16]. In Appendix I we prove the following lemmas for transformations (24a), (26a), and (26b).

Lemma 1: Let $U[0,N-1]=\mathrm{DFT}\{u[0,N-1]\}$ where the components u_k, $0\le k<N$, are real i.i.d. Gaussian random variables with mean zero and variance σ^2, and let $B_i=C_iU_i$ be the components of the sequence $B[0,N-1]$ where the C_i are complex constants and $C_i=C^*_{N-i}$, $1\le i<N$. Then application of transform (24a) to the complex subsequence $B[0,L]$ yields the real sequence $B'[0,N-1]$ whose components B'_i are independent Gaussian random variables with mean zero and variances $N\sigma^2|C_i|^2$, $0\le i<N$.

Lemma 2: Let $b[0,N-1]$ be a sequence whose components b_k, $0\le k<N$, are real i.i.d. Gaussian random variables with mean zero and variance σ^2. Then the components B'_i, $0\le i<N$, of the transform-domain sequence $B'[0,N-1]$ as obtained from (26a), are also real i.i.d. Gaussian random variables with mean zero and variance $N\sigma^2$.

Lemma 3: Let $B'[0,N-1]$ be a sequence whose components B'_i, $0\le i<N$, are real independent Gaussian random variables with mean zero and variances $N\sigma_i^2$ and where $\sigma_i=\sigma_{N-i}$, $1\le i<N$. Then the components b_k, $0\le k<N$, of the time-domain sequence $b[0,N-1]$ as obtained from (26b), are real correlated Gaussian random variables with mean zero and covariances $\tilde{r}_n$, $0\le n<N$, given by

$$\tilde{r}_n=E[b_{k+n}b_k]=N^{-1}\sum_{i=0}^{N-1}\sigma_i^2\cos(2\pi ni/N),\qquad 0\le k\le k+n<N.$$

Using Parseval's relation for the DFT [15, p. 125], one finds that the original block-energy constraint (7) becomes, in the transform domain,

$$\sum_{i=0}^{N-1} E\left[X_i'^2\right]\le N^2E_s. \tag{27}$$

By Lemma 1, the V'_i in (25) are statistically independent Gaussian random variables with mean zero and variance

$$\sigma_i^2=N(N_0/2)|\tilde{H}_i|^{-2},\qquad 0\le i<N. \tag{28}$$

Thus it follows from (25) that the equivalent transform-domain channel model for the NCGC is a set of N parallel discrete memoryless additive Gaussian noise channels where the channel inputs X'_i, $0\le i<N$, satisfy (27). This equivalence implies

$$\sup_{q_N} I(x[0,N-1];\tilde{y}[0,N-1])=\sup_{Q_N} I(X'[0,N-1];Y'[0,N-1]) \tag{29}$$

where Q_N is the class of probability densities for $X'[0,N-1]$ satisfying block-energy constraint (27). To write (29), we have made use of the fact that the average mutual information between two sequences is invariant to any succession of reversible transformations of one or both of the sequences [8, p. 30]. Application of a theorem due to Gallager [8, theorem 7.5.1] to this transform-domain channel model provides the solution for $\tilde{I}_N(E_s)$ in the form of the parametric expression

$$\tilde{I}_N(E_s) = (2N)^{-1} \sum_{i=0}^{N-1} \log\left[\max\left(\Theta|\tilde{H}_i|^2, 1\right)\right] \quad (30a)$$

where the parameter Θ is the solution of

$$\sum_{i=0}^{N-1} \max\left(\Theta - |\tilde{H}_i|^{-2}, 0\right) = 2NE_s/N_0. \quad (30b)$$

B. Properties of the NCGC

Solution (30) was obtained under the assumption that $\tilde{H}_i \neq 0$, $0 \le i < N$. However, (30a) indicates that component channel i does not contribute to $\tilde{I}_N(E_s)$ whenever $\Theta|\tilde{H}_i|^2 \le 1$, a condition which certainly holds when $\tilde{H}_i = 0$. This implies that any optimal transmission scheme will not make use of those component channels for which $\tilde{H}_i = 0$. Therefore, to include this case in the solution, we take the sum in (30b) only over those i where $\tilde{H}_i \neq 0$.

It is implied in [8, theorem 7.5.1] that $I(X'[0, N-1]; Y'[0, N-1])$ in (29) is optimized by choosing Q_N such that the transform-domain inputs X_i', $0 \le i < N$, are statistically independent Gaussian random variables with mean zero and variances $E[X_i'^2] = N\epsilon_i$, where the ϵ_i are given in (17b); i.e., (27) and thus (7) hold with equality. We note that $N\epsilon_i$ is the average input energy that must be used in the ith component channel. Since $\tilde{H}_i = \tilde{H}^*_{N-i}$, $1 \le i < N$, it follows that $\epsilon_i = \epsilon_{N-i}$, $1 \le i < N$, as can be seen from (17b).

Invoking Lemma 3 leads to the result that the optimizing q_N in (21) is such that the time-domain inputs x_k, $0 \le k < N$, are statistically correlated Gaussian random variables with mean zero and covariances $\tilde{r}_n$, $0 \le n < N$, as given in (17a). Moreover, from (17a) and with equality in (27), it follows immediately that $\tilde{r}_0 = E[x_k^2] = E_s$, $0 \le k < N$. We conclude that the optimizing q_N in (21) also satisfies the stronger constraint (13) with equality; thus this q_n is also the optimizing q_N in (14) with the implication that $\tilde{C}_N(E_s) = \tilde{I}_N(E_s)$ for this q_N. This completes the proof of Theorem 1.

IV. Relations Between Channel Models

This section is devoted to a proof of Corollary 1. Further, we also consider the case where the input symbols are *chosen* to be i.i.d. Gaussian random variables, and we show that the resulting information rate is readily obtained using the same approach as for capacity.

In Section I we stated that for the NCGC, in contrast to the DTGC and the N-block DTGC, no need exists for an initializing input sequence. At the discrete-time instant $k = 0$ (when we assume that we begin to observe the output), the initial state of the basic channel model of the NCGC is represented by the subsequence $x[N-M, N-1]$. This means that the linear convolution in (1) can be made into a circular convolution if we choose $x_k = x_{N+k}$, $-M \le k < 0$. Conversely, by letting $x_k = 0$, $N-M \le k < N$, in (11) the circular convolution becomes a linear one. In the following we shall frequently use these facts to establish fundamental relations between the three channel models introduced in Section I.

A. Proof of Corollary 1

Theorem 3: The capacities of the NCGC and channels of the N-block DTGC type are related by

$$(1-M/N)\hat{C}_{N-M}(NE_s/(N-M)) \le \tilde{C}_N(E_s) \le (1+M/N)\hat{C}_{N+M}(E_s). \quad (31)$$

Proof: The lower bound in (31) is proved first. Defining

$$\beta_{0,N-1} = \sum_{k=0}^{N-1} E\left[x_k^2\right] \quad (32a)$$

and using the result $\tilde{C}_N(E_s) = \tilde{I}_N(E_s)$ obtained in Section III, we obtain from definition (21)

$$\begin{aligned} N\tilde{C}_N(E_s) &\ge \sup_{\left\{\substack{\beta_{0,N-1} \le NE_s \\ x_k = 0,\, N-M \le k < N}\right\}} I(x[0,N-1]; \tilde{y}[0,N-1]) \\ &= \sup_{\{\beta_{0,N-M-1} \le NE_s\}} I(x[0,N-M-1]; y[0,N-1]) \\ &\ge \sup_{\{\beta_{0,N-M-1} \le NE_s\}} I(x[0,N-M-1]; y[0,N-M-1]) \\ &= (N-M)\hat{C}_{N-M}(NE_s/(N-M)) \end{aligned} \quad (32b)$$

where the suprema are over all probability densities q_N satisfying the indicated constraints. The first inequality holds because an additional input constraint can only decrease average mutual information, and the first equality holds because circular convolution becomes linear convolution for this input constraint. The second inequality holds because information can only be lost if the received sample vector is truncated [8, pp. 16–27], and the last equality follows from definition (8).

The upper bound in (31) is proved by assuming, for notational convenience, that the first input digit is transmitted at time instant $k = -M$, i.e., we let $m = -M$ in (4). Thus

$$\begin{aligned} (N+M)\hat{C}_{N+M}(E_s) &\ge (N+M)I_{N+M}(E_s) \\ &\ge \sup_{\left\{\substack{E[x_k^2] \le E_s,\, -M \le k < N \\ x_k = x_{N+k},\, -M \le k < 0}\right\}} \cdot I(x[-M,N-1]; y[-M,N-1]) \\ &\ge \sup_{\{E[x_k^2] \le E_s,\, 0 \le k < N\}} I(x[0,N-1]; \tilde{y}[0,N-1]) \\ &= N\tilde{C}_N(E_s) \end{aligned} \quad (32c)$$

where the first two inequalities hold because stronger input constraints can only decrease average mutual information. The third inequality holds because of truncation of the input and output sequences and because the constraint $x_k = x_{N+k}$, $-M \le k < 0$, implies that circular and linear convolution coincide. The last equality follows from definition (14). From Theorem 3 and definition (9), one immediately has the following corollary.

Corollary 2: The quantity $\hat{C}(E_s)$, defined as the limit as $N \to \infty$ of the capacity of the N-block DTGC $\hat{C}_N(E_s)$, may be obtained as

$$\hat{C}(E_s) = \lim_{N \to \infty} \tilde{C}_N(E_s). \tag{33}$$

Lemma 4: The optimized average block mutual information of the DTGC, defined in (6b), and the capacity of the NCGC are related by

$$(1 - M/N) I_{N-M}(NE_s/(N-M)) \le \tilde{C}_N(E_s) \le (1 + M/N) I_{N+M}(E_s). \tag{34}$$

The lower bound in (34) follows directly from (32b) and the obvious inequality $\hat{C}_N(E_s) \ge I_N(E_s)$, and the upper bound is implicit in (32c). From Lemma 4 and definition (6), one obtains the following.

Corollary 3: The capacity of the DTGC, $C(E_s)$, may be obtained as

$$C(E_s) = \lim_{N \to \infty} \tilde{C}_N(E_s). \tag{35}$$

Corollary 1 (Section II) now follows from Corollaries 2 and 3 and definition (15).

B. Information Rate for White Gaussian Inputs

A further interesting relation between channel models of the DTGC and the NCGC types may be obtained by *choosing* q_N such that the inputs x_k, $0 \le k < N$, are i.i.d. Gaussian random variables with mean zero and variance E_s. Call the resulting channel model DTGC-G and NCGC-G to distinguish them from the DTGC and the NCGC, respectively. For the DTGC-G we define the information rate

$$I^G(E_s) = \lim_{N \to \infty} I_N^G(E_s) \tag{36a}$$

where

$$I_N^G(E_s) = N^{-1} I(x[0, N-1]; y[0, N-1]) \tag{36b}$$

with $x[-M, -1]$ as in (6c). Similarly, for the NCGC-G we define

$$\tilde{I}_N^G(E_s) = N^{-1} I(x[0, N-1]; \tilde{y}[0, N-1]). \tag{37}$$

Theorem 4: The average block mutual informations of the DTGC-G and the NCGC-G (in *bits per channel input symbol* when logarithms are taken to the base 2) are related by

$$(1 - M/N) I_{N-M}^G(E_s) \le \tilde{I}_N^G(E_s) \le (1 + M/N) I_{N+M}^G(E_s) \tag{38a}$$

where

$$\tilde{I}_N^G(E_s) = (2N)^{-1} \sum_{i=0}^{N-1} \log\left[1 + 2(E_s/N_0)|\tilde{H}_i|^2\right]. \tag{38b}$$

Corollary 4: The information rate of the DTGC-G, $I^G(E_s)$, is obtained as

$$I^G(E_s) = \lim_{N \to \infty} \tilde{I}_N^G(E_s). \tag{39}$$

Theorem 4 is proved in Appendix II; Corollary 4 follows directly from (38a) and definition (36a).

V. Derivation of $C(E_s)$ and $I^G(E_s)$

In this section we prove Theorem 2 (Section II) and evaluate (39). In Section I we introduced the hypothetical NCGC which was described within an N-dimensional space. The relations derived in Section IV indicate, however, that the circularity restriction imposed becomes less and less important as N is increased. Thus the infinite-dimensional generalization of the results obtained for the NCGC (NCGC-G) will yield the corresponding results for the DTGC (DTGC-G). With respect to N, the NCGC (NCGC-G) and the DTGC (DTGC-G) are thus asymptotically equivalent.

From (16b) and the expression for the channel transfer function in (2), it follows that $|\tilde{H}_i|^2 = |H(\lambda_i)|^2$, $0 \le i < N$, where $\lambda_i = i\,\Delta\lambda_N$, $0 \le i \le L$, $\lambda_i = (i - N)\,\Delta\lambda_N$, $L < i < N$, and $\Delta\lambda_N = 2\pi/N$. We shall make use of the following simple property of Riemann integrals.

Lemma 5: Let $G(\cdot)$ be a continuous real-valued function. Then

$$\lim_{N \to \infty} N^{-1} \sum_{i=0}^{N-1} G(|\tilde{H}_i|^2) = \lim_{N \to \infty} (2\pi)^{-1} \sum_{i=0}^{N-1} G(|H(\lambda_i)|^2)\,\Delta\lambda_N = (2\pi)^{-1} \int_{-\pi}^{\pi} G(|H(\lambda)|^2)\, d\lambda. \tag{40}$$

The first part of Theorem 2 is now proved by Corollary 3 and application of Lemma 5 to the parametric expressions in (16). Using the fact that $|H(\lambda)| = |H(-\lambda)|$, $|\lambda| \le \pi$, leads to the final result (19).

Assuming that E_s is finite, it can be shown that the real covariance sequence $\tilde{r}[0, N-1]$ in (17) is positive definite [17, p. 473], thus confirming that the components of the input sequence $x[0, N-1]$ belong to a (wide-sense) stationary random process. The space over which this process is defined is N-dimensional. We may now generalize and consider the case where $N \to \infty$. The covariances $r_n = r_{-n}$, $0 \le n < \infty$, exist as the Fourier–Stieltjes coefficients of the spectral distribution function $F(\lambda)$, $|\lambda| \le \pi$, in the form

$$r_n = E[x_{k+n} x_k] = \int_{-\pi}^{\pi} e^{jn\lambda}\, dF(\lambda), \qquad -\infty < k < \infty, \tag{41}$$

provided that $F(\lambda) \ge 0$ is absolutely continuous and nondecreasing in the interval $|\lambda| \leq \pi$, $F(-\pi)=0$, and $F(\pi) = r_0 < \infty$ [17, pp. 474–476], [18, pp. 8–18]. On the other hand, (17) and Lemma 5, with $\epsilon_i = \epsilon_{N-i}$, $1 \leq i < N$, give

$$\lim_{N \to \infty} \tilde{r}_n = (2\pi)^{-1} \int_{-\pi}^{\pi} S_x(\lambda) e^{jn\lambda} \, d\lambda \tag{42}$$

where $S_x(\lambda)$ is given in (20b). For the relation

$$r_n = \lim_{N \to \infty} \tilde{r}_n \tag{43}$$

to hold, it follows by comparison of (41) and (42) that $F(\lambda)$ must be of the form

$$F(\lambda) = (2\pi)^{-1} \int_{-\pi}^{\lambda} S_x(\nu) \, d\nu, \qquad |\lambda| \leq \pi. \tag{44}$$

The right side of (44) represents $F(\lambda)$ as required by (41). In particular, $F(\lambda) \geq 0$, $F(-\pi)=0$, and $F(\lambda)$ is nondecreasing since $S_x(\lambda) \geq 0$. Furthermore, using $\tilde{r}_0 = E_s$ (Theorem 1), it follows from (42) and (44) that

$$F(\pi) = \lim_{N \to \infty} \tilde{r}_0 = E_s. \tag{45}$$

Therefore, (43) is valid, and we obtain (20a) from (42) using the fact that $S_x(\lambda)$ is an even function. It follows that the capacity-achieving input process of the DTGC, $\{x_k\}$, $-\infty < k < \infty$, is zero-mean Gaussian having a continuous spectral distribution function $F(\lambda)$ with corresponding spectral density $S_x(\lambda)$. As a consequence, we obtain the result that capacity is achieved when all inputs x_k, $-\infty < k < \infty$, have the same average energy E_s.

From (20b), we note that (19) has the well-known water-filling interpretation of capacity [8, p. 389]. For $C(E_s)$ to be finite, a constraint must be placed on $H(\lambda)$. Assuming $0 < \Theta < \infty$ and using $\ln[\max(z,1)] \leq \ln(1+z) \leq z$, $z \geq 0$, (19a) yields the sufficient condition

$$\int_0^{\pi} |H(\lambda)|^2 \, d\lambda < \infty \tag{46}$$

for the capacity to be finite, which is equivalent to stating that the unit-sample response of the channel filter, $h[0, M]$, has finite energy. This completes the proof of Theorem 2. Finally, we invoke Lemma 5 to obtain $I^G(E_s)$ from (38b) and Corollary 4 as

$$I^G(E_s) = (2\pi)^{-1} \int_0^{\pi} \log\left[1 + 2(E_s/N_0)|H(\lambda)|^2\right] d\lambda. \tag{47}$$

VI. Summary and Conclusion

The capacity of the discrete-time Gaussian channel with ISI, where the inputs are subject to a *per symbol* average-energy constraint (DTGC), has been derived. The result was obtained through a novel indirect approach by introducing first the concept of the N-circular Gaussian channel (NCGC) whose capacity is readily derived using properties of the discrete Fourier transform (DFT). The capacity of the DTGC was then found with the help of simple relations which were proved to hold between the different channel models. Studying first the hypothetical NCGC provides a better understanding of the DTGC. For the channel where the inputs are subject to a *per block* average-energy constraint (N-block DTGC), capacity results were previously obtained using asymptotic properties of Toeplitz forms [9]–[13]. However, the validity of Corollary 1 was (to our knowledge) never formally proved. Moreover, the capacity of the NCGC, $\tilde{C}_N(E_s)$, can be used to approximate the capacity of the DTGC, $C(E_s)$, to any desired degree of accuracy as N increases. This may be of some interest when $C(E_s)$ is to be evaluated numerically.

As an additional result, we obtained the solution for the information rate of the DTGC-G, $I^G(E_s)$, where the inputs are chosen to be i.i.d. zero-mean Gaussian random variables with fixed average energy E_s. This result was also obtained without resorting to methods which require the theory of asymptotic eigenvalue distribution of certain Toeplitz covariance matrices.

Finally, note that our method is not applicable to the case where the channel memory is unbounded ($M = \infty$). It should be straightforward, however, to extend our approach to the more general case of the discrete-time Gaussian (vector) channel with multiple inputs and outputs [10], [12], [13] provided that the channel memory is finite.

Appendix I

Proof of Lemma 1: We want to show that the B_i', $0 \leq i < N$, are Gaussian random variables with

$$E[B_i'] = 0 \quad \text{and} \quad E[B_i' B_k'] = N\sigma^2 |C_i|^2 \delta_{i-k}, \qquad 0 \leq k < N. \tag{I.1}$$

It follows from (22a) that the complex $B_i = C_i U_i$, $0 \leq i < N$, are a weighted sum of the u_k, $0 \leq k < N$, which are (by definition) real i.i.d. zero-mean Gaussian random variables. Therefore, the B_i' obtained from (24a) are also zero-mean Gaussian. To prove the covariances in (I.1), it is required to evaluate the expectations $E[B_i^R B_k^R]$, $E[B_i^R B_k^I]$, and $E[B_i^I B_k^I]$, $0 \leq i \leq L$, $0 \leq k \leq L$. We first form the products $B_i^R B_k^R$, $B_i^R B_k^I$, and $B_i^I B_k^I$ by substituting $B_i^R = (B_i + B_i^*)/2$ and $B_i^I = -j(B_i - B_i^*)/2$ to obtain

$$B_i^R B_k^R = (B_i B_k + B_i B_k^* + B_i^* B_k + B_i^* B_k^*)/4 \tag{I.2a}$$

$$B_i^R B_k^I = -j(B_i B_k - B_i B_k^* + B_i^* B_k - B_i^* B_k^*)/4 \tag{I.2b}$$

$$B_i^I B_k^I = -(B_i B_k - B_i B_k^* - B_i^* B_k + B_i^* B_k^*)/4. \tag{I.2c}$$

Since $E[B_i^* B_k^*] = (E[B_i B_k])^*$ and $E[B_i^* B_k] = (E[B_i B_k^*])^*$, the expectations of the products in (I.2) can be expressed in terms of $E[B_i B_k]$ and $E[B_i B_k^*]$ where

$$E[B_i B_k] = (C_i C_k) E[U_i U_k] = N\sigma^2 |C_i|^2 \delta_{i-k}, \qquad i = \begin{cases} 0, & N \text{ odd} \\ 0 \text{ or } L, & N \text{ even} \end{cases} \tag{I.3a}$$

and

$$E[B_i B_k^*] = (C_i C_k^*) E[U_i U_k^*] = N\sigma^2 |C_i|^2 \delta_{i-k}, \qquad 0 \leq i \leq L. \tag{I.3b}$$

In (I.3), we have used $E[u_m u_n] = \sigma^2 \delta_{m-n}$ to obtain

$$E[U_i U_k] = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} E[u_m u_n] \Omega_N^{-im-kn} = \sigma^2 \sum_{m=0}^{N-1} \Omega_N^{-m(i+k)} = \begin{cases} N\sigma^2, & i+k = 0 \bmod N \\ 0, & \text{otherwise} \end{cases} \tag{I.4a}$$

and

$$E[U_i U_k^*] = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} E[u_m u_n] \Omega_N^{-im+kn}$$
$$= \sigma^2 \sum_{m=0}^{N-1} \Omega_N^{-m(i-k)} = N\sigma^2 \delta_{i-k}, \qquad 0 \le i \le L. \quad \text{(I.4b)}$$

Equation (I.4) is a consequence of the orthogonality relationships between powers of the exponential Ω_N [15, p. 88]. Taking expectation on both sides of (I.2) and using (I.3) gives

$$E[B_i^R B_k^R] = \begin{cases} N\sigma^2 |C_i|^2 \delta_{i-k}, & i = \begin{cases} 0, & N \text{ odd} \\ 0 \text{ or } L, & N \text{ even} \end{cases} \\ (N/2)\sigma^2 |C_i|^2 \delta_{i-k}, & 1 \le i \le \begin{cases} L, & N \text{ odd} \\ L-1, & N \text{ even} \end{cases} \end{cases} \quad \text{(I.5a)}$$

$$E[B_i^I B_k^I] = (N/2)\sigma^2 |C_i|^2 \delta_{i-k}, \qquad 1 \le i \le \begin{cases} L, & N \text{ odd} \\ L-1, & N \text{ even} \end{cases} \quad \text{(I.5b)}$$

$$E[B_i^R B_k^I] = 0, \qquad 0 \le i \le L, \quad 0 \le k \le L. \quad \text{(I.5c)}$$

From (24a) and (I.5), it follows that the covariances $E[B_i' B_k']$ are indeed as given in (I.1). The B_i', $0 \le i < N$, are therefore uncorrelated, and since they are Gaussian, they are also independent.

Proof of Lemma 2: Since transformation (26) is equivalent to the combination of transforms (22) and (24), Lemma 2 is a special case of Lemma 1. The proof of Lemma 2 is obtained by replacing $u[0, N-1]$ and $U[0, N-1]$ in the proof of Lemma 1 by $b[0, N-1]$ and $B[0, N-1]$, respectively, and letting $C_i = 1$, $0 \le i < N$.

Proof of Lemma 3: By definition, the B_i', $0 \le i < N$, are statistically independent Gaussian random variables with

$$E[B_i'] = 0 \quad \text{and} \quad E[B_i' B_k'] = N\sigma_i^2 \delta_{i-k}. \quad \text{(I.6)}$$

According to (26b), the b_k, $0 \le k < N$, are obtained from a weighted sum of zero-mean Gaussian random variables; therefore, the b_k are also zero-mean Gaussian. The covariances are obtained from (26b) as

$$E[b_k b_m] = N^{-2} \Big\{ \sum_{i=0}^{L} \sum_{n=0}^{L} d_i d_n E[B_i' B_n'] \cdot \cos(2\pi ki/N)\cos(2\pi mn/N)$$
$$+ \sum_{i=0}^{L} \sum_{n=L+1}^{N-1} d_i d_n E[B_i' B_n'] \cdot \cos(2\pi ki/N)\sin(2\pi mn/N)$$
$$+ \sum_{i=L+1}^{N-1} \sum_{n=0}^{L} d_i d_n E[B_i' B_n'] \cdot \sin(2\pi ki/N)\cos(2\pi mn/N)$$
$$+ \sum_{i=L+1}^{N-1} \sum_{n=L+1}^{N-1} d_i d_n E[B_i' B_n'] \cdot \sin(2\pi ki/N)\sin(2\pi mn/N) \Big\}$$
$$= N^{-1} \Big\{ \sum_{i=0}^{L} d_i^2 \sigma_i^2 \cos(2\pi ki/N)\cos(2\pi mi/N) + \sum_{i=L+1}^{N-1} d_i^2 \sigma_i^2 \sin(2\pi ki/N)\sin(2\pi mi/N) \Big\}. \quad \text{(I.7)}$$

For N odd, we find from (I.7) and (26c) with $\sigma_i = \sigma_{N-i}$, $1 \le i < N$,

$$E[b_k b_m] = N^{-1} \Big\{ \sigma_0^2 + 2\sum_{i=1}^{L} \sigma_i^2 [\cos(2\pi ki/N)\cos(2\pi mi/N) + \sin(2\pi ki/N)\sin(2\pi mi/N)] \Big\}$$
$$= N^{-1} \Big\{ \sigma_0^2 + 2\sum_{i=1}^{L} \sigma_i^2 \cos[2\pi(k-m)i/N] \Big\}$$
$$= N^{-1} \sum_{i=0}^{N-1} \sigma_i^2 \cos[2\pi(k-m)i/N],$$
$$0 \le k < N, 0 \le m < N. \quad \text{(I.8)}$$

For N even the last equality is obtained similarly. The covariance matrix is completely specified by N different covariances since $E[b_k b_m]$ depends only on the absolute (time) difference $|k-m|$ (the covariance matrix is circular, i.e., its first row specifies it). Letting $m = k + n$ in (I.8) and defining $\tilde{r}_n = E[b_{k+n} b_k]$, $0 \le n < N$, completes the proof of Lemma 3.

Appendix II

Proof of Theorem 4: From definitions (37) and (36b), we obtain the lower bound in (38a) using arguments similar to those used in the development of (31). Thus

$$N\tilde{I}_N^G(E_s) \ge \{ I(x[0, N-1]; \tilde{y}[0, N-1]) : x_k = 0, N-M \le k < N \}$$
$$= I(x[0, N-M-1]; y[0, N-1])$$
$$\ge I(x[0, N-M-1]; y[0, N-M-1])$$
$$= (N-M) I_{N-M}^G(E_s). \quad \text{(II.1a)}$$

The upper bound in (38a) is proved similarly by using (4) with $m = -M$. Thus

$$(N+M) I_{N+M}^G(E_s) \ge \{ I(x[-M, N-1]; y[-M, N-1]) : x_k = x_{N+k}, -M \le k < 0 \}$$
$$\ge \{ I(x[0, N-1]; y[0, N-1]) : x_k = x_{N+k}, -M \le k < 0 \}$$
$$= I(x[0, N-1]; \tilde{y}[0, N-1])$$
$$= N\tilde{I}_N^G(E_s). \quad \text{(II.1b)}$$

To prove (38b), we invoke Lemma 2 to show that the transform-domain inputs X_i', $0 \le i < N$, are i.i.d. Gaussian random variables with mean zero and variance NE_s. The equivalent transform-domain channel model is thus a set of N parallel, but independent, memoryless additive Gaussian noise channels each having zero-mean Gaussian inputs of variance NE_s. The additive noise in the ith channel is independent and zero-mean Gaussian with variance $N(N_0/2)|\tilde{H}_i|^{-2}$, as previously shown in proving (28). The average mutual information between input and output of the ith component channel is then simply given by [8, p. 32]

$$I(X_i'; Y_i') = (1/2)\log[1 + 2(E_s/N_0)|\tilde{H}_i|^2], \quad \text{(II.2)}$$

and (38b) follows immediately because the N component channels are mutually independent. Clearly, component channels where $\tilde{H}_i = 0$ do not contribute to $\tilde{I}_N^G(E_s)$. This completes the proof of Theorem 4.

References

[1] R. E. Blahut, "Capacity of a hard-limited bandpass Gaussian channel," in *Abstracts of Papers, IEEE Int. Symp. Inform. Theory*, Brighton, England, June 1985, p. 122.

[2] A. R. Calderbank, L. H. Ozarow, and C. Heegard, "A new approach to high-density magnetic recording," in *Abstracts of Papers, IEEE Int. Symp. Inform. Theory*, Brighton, England, June 1985, p. 122.

[3] J. K. Wolf and G. Ungerboeck, "Trellis coding for partial-response channels," *IEEE Trans. Commun.*, vol. COM-34, pp. 765–773, Aug. 1986.

[4] R. W. Lucky, J. Salz, and E. J. Weldon, *Principles of Data Communication*. New York: McGraw-Hill, 1968.

[5] P. Kabal and S. Pasupathy, "Partial-response signaling," *IEEE Trans. Commun.*, vol. COM-23, pp. 921–934, Sept. 1975.

[6] G. D. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 363–378, May 1972.

[7] I. N. Andersen, "Sample-whitened matched filters," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 653–660, Sept. 1973.

[8] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[9] B. S. Tsybakov, "Capacity of a discrete-time Gaussian channel with a filter," *Probl. Peredach. Inform.*, vol. 6, pp. 78–82, 1970.

[10] ——, "Capacity of a vector Gaussian channel without memory," *Probl. Peredach. Inform.*, vol. 1, pp. 26–40, 1965.

[11] R. M. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 725–730, Nov. 1972.

[12] W. Toms and T. Berger, "Capacity and error exponents of a channel modeled as a linear dynamic system," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 124–126, Jan. 1973.

[13] L. H. Brandenburg and A. D. Wyner, "Capacity of the Guassian channel with memory: The multivariate case," *Bell Syst. Tech. J.*, vol. 53, pp. 745–778, May–June 1974.

[14] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*. Berkeley, CA: University of California Press, 1958.

[15] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.

[16] O. Ersoy, "Real discrete Fourier transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 880–882, Aug. 1985.

[17] J. L. Doob, *Stochastic Processes*. New York: Wiley, 1953.

[18] E. J. Hannan, *Time Series Analysis*. London: Science Paperbacks and Methuen, 1967.